ARPA Network Information Center
Stanford Research Institute
Menlo Park, California 94025

NIC #
Packet Radio Temporary Note 19

R. L. Kaiser (Collins)
22 January 1973

SPREAD-SPECTRUM CONSIDERATIONS

1.0 INTRODUCTION

The term spread spectrum implies a modulation technique which spreads the signal being conveyed over a wider bandwidth than would be required to convey the basic message content of the signal. This, in turn, implies wasted bandwidth and/or energy with respect to the total message-bearing capacity of the transmission media. Such wasted capacity can only be justified if an offsetting advantage is obtained through the use of spread-spectrum techniques. For the packet message network, the advantage which may be gained through the use of spread-spectrum techniques is the reduction of interference between two or more overlapping message packets. This, in turn, minimizes the requirement for retransmission with its attendent message delays and wasted capacity of the transmission media. This note provides a brief review of some of the aspects of spread-spectrum techniques.

2.0 PRINCIPLES OF SPREAD SPECTRUM

The basic concepts underlying all spread-spectrum techniques are those which are embodied in communication theory with respect to matched filters and correlation detection. That is, when the receiver has exact knowledge of the possible waveforms which it will be receiving, then the receiver can be configured to select the desired set of waveforms and exclude the other waveforms which appear simultaneously at the receiver input. The degree of selection and exclusion depends upon two factors:

a. The degree to which the receiver appears as a perfect matched filter to the desired waveforms
b. The degree to which the desired waveforms and the undesired waveforms are orthogonal (uncorrelated).

In general, perfect-match filtering cannot be achieved because of the effects of distortion by the transmission medium. Similarly, perfect orthogonality cannot be achieved because of timing errors. Further, it can be shown that the number of orthogonal waveform/codes which can be employed is directly proportional to the transmission bandwidth. In general, the larger the transmission bandwidth, the greater the ability to reject interfering signals. The validity of the last statement is dependent on the linearity of the transmission medium.

For example, it can be shown that for radio transmission path subject to multipath, there is an optimum spread-spectrum bandwidth for each link and the use of excess bandwidths can result in reduction of signal quality.

If the interfering signal is noncorrelated with the desired signal, then the output signal-to-noise ratio is given by the following expression:

$$SNR_o = \frac{S}{S_i + N} \left(\frac{W}{b}\right)$$

where:

S is the desired signal strength

S_i is the interfering signal strength in the transmission bandwidth, W

N is the noise level in the transmission bandwidth, W

b is the information bandwidth.

Most spread-spectrum techniques employ digital modulation techniques with chip rates which are an order of magnitude or greater than the basic message data rate. From the above formula it is seen that if W = 10b, S/N = 10, and S_i = S, that SNR_o = 9, compared to a value of 100 in the absence of no interference or to a value of 0.9 in the presence of interference without spread-spectrum techniques.

One of the detection approaches commonly employed is to cross-correlate the received signal with a stored replica of the desired spreading waveform/code and then process the product through integrate-and-dump filters. The cross-correlation process will despread the desired signal; that is, it will coherently collapse the spread signal in a bandwidth, W, into the original message signal in a bandwidth, b. Any interfering signal simply becomes respread and if the interfering signal is uncorrelated, simply appears as an increase in the background noise level. The integrate-and-dump filter simply provides an output which is used to decide whether or not the output of the crosscorrelation is a desired signal or not.

The decision process involves a minimum time delay on the order of $T = \frac{1}{b}$ second, assuming that the optimum integration time is employed; that is, the maximum output of the integrate-and-dump filter reaches its maximum output for a correlation signal at the T seconds following the arrival of a packet.

The width of the autocorrelation output function is $\frac{1}{W}$ second, and it is necessary to synchronize the integrate-and-dump filter with the time of the packet arrival within $\pm\frac{1}{2W}$ second.

The latter fact points out that the greater the spread bandwidth, the greater the synchronization problem. It also indicates a limitation on discrimination between signals according to time of arrival. That is, if an overlapping signal with the same spread waveform arrives within $\frac{1}{W}$ second after the arrival of the initial signal, then the receiver will not be able to reject the second interfering signal. In practice, hardware limitations may require an even greater separation in arrival times in order to ensure rejection of the late arriving signal. The implications of the use of greater bandwidths for spread-spectrum techniques can be summarized as follows:

Advantages of Greater Bandwidth	Disadvantages of Greater Bandwidth
1. Greater ability to reject interference	1. Wasted capacity of transmission medium
2. Greater number of orthogonal waveform/codes.	2. Greater costs in equipment complexity.

3.0 PSEUDORANDOM SEQUENCES

At the heart of most spread-spectrum systems is a pseudorandom sequence generator. There are several reasons why this is so. For the application of interest in this paper, the primary reason is their simplicity of generating orthogonal codes, their flexibility, and their positive reproducibility for reference/correlation purposes at the receiver.

Feedback shift-register generators are available that can be quickly modified by appropriate switching, to provide entirely different (ie, uncorrelated) sets of pseudorandom waveforms. Moveover, any number of units can be produced to accurately and positively generate identical sequences so that multiple-user applications can be implemented. Lastly, such pseudorandom sequences can have desirable randomness properties. The so-called maximal-length sequences have an autocorrelation function which has a tall central peak in the vicinity of zero shift, but which is very small for time shifts greater than the width of one digit. Also, the crosscorrelation function for two different maximal-length sequences of the same length is very nearly zero for all time shifts. Such sequences can be narrow-band filtered (the filter cutoff frequency much less than f_c, the digit rate), or a narrow slice taken out of the spectrum to produce a pseudonoise waveform for applications requiring "noise-like" signals. Pseudonoise generated in this fashion has the advantageous property over truly random noise in its ability to be exactly reproduced at remote locations.

The type of binary code or pseudorandom sequence under consideration in this section is that usually produced by a feedback shift-register generator or a tapped delay line. A simple shift-register consists of several (say, n) 2-state active devices (commonly referred to as flip-flops) connected in a cascade arrangement. The binary state of each stage of the register represents a binary digit or bit (0 or 1), and the "state" of the entire shift-register at any instant is characterized by the n-bit binary number formed by the n digits stored in the register. A shift register transfers the binary state (ie, the binary digit stored in each stage) to the next succeeding stage whenever a shift pulse occurs. Such shift pulses are normally produced by a "clock" which is essentially a highly stable oscillator or frequency generator. When the shift pulses are regularly spaced, as from a clock, each particular digit will propagate at a uniform rate down the shift register, much as a pulse flows down a delay line. However, we are interested here not in simple shift registers, but in feedback shift-register generators (FSRG). An FSRG is formed, in general, by taking an output from each of the n stages of the register and combining these outputs in a connection matrix and feeding the resultant output back to the input of the register. Thus, binary digits flowing out of the FSRG at the clock frequency are some function of the n preceding digits that were stored in the register. The code or pseudorandom sequence generated by an FSRG is a specific function of the interconnections between the stages of the register. Particular codes are selected by choosing certain stages for feedback and by the feedback logic used in the connection matrix. The "key" to a pseudorandom sequence usually refers to the choice of stages for inputs to the connection matrix, whereas the class of codes (ie, either linear or nonlinear) produced by the FSRG is determined determined by the type of operations performed in the connection matrix. The key may also refer to the binary coefficients, a_i = (0 or 1), in a transportation $F(a_1b_1,a_2,b_2 \ldots a_n,b_n)$, perhaps nonlinear, performed on the outputs of the n stages of the FSRG to form a new output sequence. "Linear logic" results in linear codes and higher order operations produce nonlinear codes.

The operation of an FSRG also may be visualized as a sequence of binary states, each of which is characterized by the n-bit (-digit) binary number stored in the register. The sequence of binary numbers formed in the register is a specific function of the initial state conditions of the register and of the feedback logic. Thus, given an n-bit number in the register, the FSRG produces a

new deterministic n-bit number when triggered by a shift pulse. Consequently, such a generator is periodic since the deterministic process will eventually reestablish the initial state conditions in the register; ie, the original binary number will be re-formed, and the sequence proceeds to repeat itself at that point. The length of the sequence or code is the number of digits in one period. Since the total number of n-bit binary numbers that can be formed from n binary digits is 2^n, and the binary code produced by the generator is a specified function of the n-bits in the register, the maximum length of such a code is 2^n digits and the maximum period is $2^n/f_c$ second, where f_c is the clock frequency. If the period of the code is long enough, the code has the appearance of randomness over any subsection of the period, despite the deterministic method of generation, and it is termed a pseudorandom sequence.

The most general code produced by an FSRG is composed of binary digits, b_i, either 0 or 1, determined mathematically by a module-2 recursion relation of the form

$$b_i = \sum_{\text{mod-2}} [c_o + c_j b_{i-j} + c_{jk}\, b_{i-j} b_{i-k} + c_{jk\ell}\, b_{i-j} b_{i-k}\, b_{i-\ell} \ldots] \qquad (3\text{-}1)$$

where:

b_{i-j} is the j^{th} binary digit in the register preceding b_i .

The c-s are binary coefficients, either 0 or 1, that describe the operations performed in the connection matrix. The notation of (3-1) indicates that the summation is extended over all possible single terms, pairs, triples, etc of the n digits preceding b_i. Also, the summation is performed by module-2 or mod-2 addition, defined by the rules 1 + 1 = 0 + 0 = 0 and 1 + 0 = 0 + 1 = 1. A module-2 addition device simply compares two inputs and generates the binary symbol "1" if the inputs are dissimilar and a "0" if they are alike. The total number, N_c, of possible coefficients in (3-1) is the total number of combinations of n things taken none, one, two, ... up to n at a time, or

$$N_c = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} = (1+1)^n = 2^n \quad . \qquad (3\text{-}2)$$

Thus, the total number of possible coefficients is equal to the maximum possible number of code digits in one period of the code. Since the 2^n coefficients may be chosen arbitrarily, it appears that an FSRG implementing equation (3-1) can

Digitized by the Internet Archive
in 2025 with funding from
Amateur Radio Digital Communications, Grant 151

https://archive.org/details/temporary-note-019

produce an arbitrary code. However, in general, a given set of feedback connections produces sequences of periods less than 2^n digits, since repetition commences at any point where an n-digit binary number is repeated in the register. The complete set of distinct periodic codes produced by any particular set of feedback connections will total 2^n, the difference codes being generated with different initial state conditions in the register.

The simplest type of feedback connection that may be used is one retaining only the linear terms in (3-1). The appropriate formula is a special case of the general form as follows:

$$b_i = c_o + \sum_{j=1}^{n} c_j b_{i-j} \qquad (3\text{-}3)$$

where:

all digits are binary, either 0 or 1, and the addition is mod-2.

Whenever the digits of a sequence are generated by some linear (mod-2 sum) combination of the n preceding digits, the sequence is termed linear. Linear codes are of considerable importance for the reason that they can display correlation properties and spectral distributions approaching the ideal for use in producing spread-spectrum signals. For the linear case of the binary coefficients, c_j, are represented by open or closed switches in the feedback network. A closed switch causes the corresponding coefficient in (3-3) to have the value unity, whereas an open switch assigns the value zero to the coefficient.

This simple form of an FSRG is only one method of generating a linear sequence. Other more complicated methods exist. Compound FSRG's or combinations of more than one simple FSRG can also be used to produce linear sequences. Examples of these are the multiple-return generator consisting of several feedback loops around more than one simple FSRG or several simple FSRG's acting independently in a cascade arrangement, one serving as the input for the next, or connected in a "beat" or parallel fashion so that their outputs add mod-2. However, it can be shown that the sequences from these compound generators can be generated by a simple FSRG of equal or fewer total stages. Also, although we emphasized the use of flip-flop shift registers, tapped delay lines can also be used to drive the feedback logic. In any case, the sequence generated by any of these generators will be linear if mod-2 addition is the only logic operation employed. Two types of linear generators will be discussed. Generators which,

because of a particular feedback logic, produce only cyclic permutations of one sequence no matter what the initial register contents (so long as they were not all zero) are called maximal generators. Generators that produce more than one sequence for a particular logic are called nonmaximal.

The mathematical theory of linear sequences and generators has been thoroughly studied and expounded upon by several groups to the extent that we feel it unnecessary to provide proof here. However, we will distringuish between maximal and nonmaximal generators. Determination of whether a specific FSRG is maximal or nonmaximal need not be a complicated task since tables exist listing all maximals with 19 or fewer stages.

Properties of Maximal Generators and Maximal-Length Sequences

a. A maximal generator of n stages produces a maximal-length sequence of length $L = 2^n - 1$ digits. The period of such a code is L/f_c second, where f_c is the clock frequency. For example, if $n = 32$ and $f_c = 10$ kHz, the period of the sequence is almost 5 days.
b. The number of distinct maximal-length sequences for any n is $\phi(2^n - 1)/n$, where $\phi(x)$, the Euler phi-function, is the number of integers less than x which are prime to x. This number is roughly L/n.
c. The autocorrelation function of a maximal-length sequence is periodic with a tall triangular peak of height, L, and base the width of 2 digits reoccurring at multiples of the period L/f_c, and otherwise has the value -1. Since the peak value , $L = 2^n - 1$, the excellent approximation for large n, the height of the peak relative to the subsidiary off-peak autocorrelation (the pedestal) doubles per each stage of the register.
d. The power density spectrum of the maximal-length sequence (ie, the Fourier transform of the autocorrelation function) contains harmonic lines (impulses) at multiples of f_c/L with an "envelope" given by $(\sin \pi f/f_c)^2/(\pi f/f_c)^2$, which is determined solely by the triangular peaks of the autocorrelation function. Note that the bandwidth of the spectrum is proportional to the clock frequency, f_c.
e. The crosscorrelation properties between maximal-length sequences from differently connected generators with equal number of stages has been studied experimentally. It has been experimentally demonstrated that the crosscorrelation of two different maximal-length sequences of the same length is negligible. Similarly, it has been experimentally determined that the

crosscorrelation function for any time shift has roughly a Gaussian distribution of amplitudes about (ie, perpendicular to and centered on) the time axis with an average absolute value proportional to $1/\sqrt{L}$.

The above correlation and spectral properties are obtained for averages over a full period of the sequence, but for long sequences such averaging may be impractical and short-time averages would be more appropriate. Unfortunately, not a great deal is known about the short-time properties of these sequences.

Properties of Nonmaximal Generators and Their Sequences

a. The longest possible nonmaximal-length sequence (less than $2^n - 1$ digits) is produced by the initial state conditions of all zeros stored in the register except a single one in either the first or last stage. All other initial conditions produce shorter length sequences.
b. Ignoring the trivial sequence of all zeros, the sum of the lengths of all the different nonmaximal-length sequences is $2^n - 1$ digits.
c. The autocorrelation and crosscorrelation characteristics of certain nonmaximal codes have been shown to be excellent. For the longest sequences, the height of the central autocorrelation peak relative to the highest subsidiary peak doubles for every stage of a generator constructed with the minimum number of stages required to generate the sequence.
d. The autocorrelations and power spectra can be obtained by determining the equivalent compound or beat-type generator which consists of a combination of only maximal generators.

Since nonmaximal-length sequences can be generated by a combination of maximal generators, an appropriate choice of the maximal sequences used in these configurations can produce a nonmaximal sequence which has a length nearly equal to the maximal sequence length that could have been generated with the same number of total stages. That is, two generators of n_1 and n_2 stages used in combination have properties which are nearly as good as a generator of $n_1 + n_2$ stages with a maximal length $2_{n1+n2} - 1$ digits. For example, consider a 36-stage generator. In order to set up a maximal 36-stage FSRG, a knowledge of which feedback connections yield maximal generators is required, and as was mentioned before, only maximal generators of 19 or less stages have been tabulated. Thus, since one does not know how to connect up a 36-stage maximal generator, one must be satisfied with a nonmaximal-length sequence generated by

the combination of two maximal generators of, say, 19 and 17 stages. Since the periods of the two component maximal-length sequences, $L_1 = 2^{19} - 1$ and $L_2 = 2^{17} - 1$, are relatively prime (no common factors), the longest period of a nonmaximal-length sequence is $L_1 + L_2$. This length differs by less than 0.001 percent from $L = 2^{36} - 1$, the period that would have been obtained with a maximal generator of 36 stages. The long-time average secondary autocorrelation peaks are down from the central peak by a factor approximately 10^{-5}, as compared to about 10^{-11} for the maximal-length sequence. Although nonmaximal sequences are not so amenable to general analysis, they may be very close to being equally good theoretically and possibly better from a practical viewpoint, than maximal sequences. This points up the fact that nonmaximal sequences are not to be considered unimportant.

4.0 ALTERNATES TO PSEUDORANDOM SEQUENCES

There are a number of other spreading techniques which could be considered other than the use of digital pseudorandom sequences. For example, conventional frequency modulation is a spread-spectrum technique which provides rejection of interference at the expense of greater bandwidths. As long as the overlapping signal is less than 3 dB it will reject an identical frequency-modulated signal. Sometimes satellite communications multichannel-per-carrier fm systems employ additional spreading waveforms in order to spread the energy more evenly over the transmission bandwidth and to minimize interference with other satellite and terrestrial systems.

Although the randomness aspects of pseudorandom sequences were not particularly emphasized in the preceding section, this is a desired characteristic for military and secure communication links. Properly employed, the transmitted energy tends to be distributed, on the average, evenly over the transmission band.

Conceptually, one could achieve this randomness for spreading the transmitted energy evenly over the transmission bands by other means. For example, the energy could be spread by simply modulating with white Gaussian noise. The difficulty is in reproducing the noise waveform at the receive terminal for despreading. Conceptually, one could achieve the same effect by employing a finite length of white noise, retaining only the zero crossings, and matching this to a series of Walsh functions. This could easily be reproduced at the

receive terminal for despreading with the "key" being the set of coefficients which specifies the required Walsh functions.

For the packet message network, randomness to the degree often sought for military and secure links may not be as desirable. An excellent alternative could be to use single Walsh functions as the spreading code/waveform. Walsh functions are orthogonal and are easily generated/processed by digital techniques. These and related functions are being employed in a number of communication applications such as for compressed digital voice transmission, multiplexing, etc.[1]

5.0 ISSUES TO BE CONSIDERED

There are two basic issues to be considered before the suitability of the use of spread spectrum for resolving packet conflicts can be determined. One is the issue of how much interference protection is required. This involves determining a number of sub issues such as:

a. The probability of the occurrence of packet overlapping
b. How many errors per packet are acceptable?
c. How many packet retransmissions are acceptable?

Once it is determined how much interference protection is required, then the issue of what is the most cost-effective technique for achieving the required protection can be properly considered.

The answer to the questions associated with the issue of the degree of protection required will, in part, provide overall direction for evaluation of candidate approaches for resolving conflicts when packets overlap. The degree of protection will properly be dependent upon the nature of the message being conveyed by the packet. For example, nonpriority compressed voice can tolerate a greater number of errors per packet, as well as a greater number of retransmissions, than can be tolerated by priority computer-to-computer message packets.

It may develop that an acceptable level of protection and packet retransmissions can be obtained with just a few orthogonal spread-spectrum waveforms/codes, far less than the number of terminals. Thus, it may be possible to consider a plan whereby only a few orthogonal waveforms/codes are employed and these, in turn,

---

[1] Lackey, R.B., "The Wonderful World of Walsh Functions," ICC 1972, Paper 38-1, June 1972.

are distributed to the message origination station randomly on the basis of one per origination station. Each receiving station would be equipped to provide a match for all of the waveform/codes in use. In the event of packet overlap, the receiver could behave in several possible ways. One way would be to give precedence to the signal which arrived first and to reject the others. This would require a retransmission of those which had been rejected. Alternatively, the receiver could be equipped to separate and process simultaneously as many different packets as there are different waveform/codes. In the latter case, the overlap of two packets with the same waveform/code would result in rejection of the one which arrived late. The choice between the two approaches depends upon the difference in receiver complexity and cost as well as the allowable number of retransmissions.

There are other possible uses of spread spectrum in the packet message network than resolving conflict between overlapping packets. For example, there is the possiblity of using spread-spectrum modulations to increase multipath discrimination. Spread spectrum could also be used to insert into the network, in a noninterfering manner, signals for test, synchronization, bookkeeping, or emergency channel purposes. These signals could be inserted as low-level (say, 20 dB down) spread-spectrum signals. Examples of the use of this mode would be to check path conditions, position reporting for mobile terminals, repeater diagnosis, emergency channels under severe transmission conditions, a broadcast mode to overcome repeater outages, and/or as a reservation channel.

In summary, the following issues need to be considered:

a. How much protection required for overlapping packets?
b. What would be the spread-spectrum characteristics necessary to provide such protection?
c. What are merits of different coding/waveforms?
    1. Pseudorandom maximal
    2. Pseudorandom nonmaximal
    3. Walsh functions
    4. Frequency modulation
d. Hardware technology for hand-held terminal implementation
    1. Card storage requirements
    2. Logic elements